Package 'ActivePathways'

May 4, 2019

Title Multivariate Pathway Enrichment Analysis

Version 1.0.1

Description An integrative method of analyzing multi omics data that conducts enrichment analysis of annotated gene sets. 'ActivePathways' uses a statistical data fusion approach, rationalizes contributing evidence and highlights associated genes, improving systems-level understanding of cellular organization in health and disease.

Depends R (>= 3.2.3)

Imports metap, data.table, ggplot2

License GPL-3

BugReports https://github.com/reimandlab/ActivePathways/issues

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Suggests testthat, knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Helen Zhu [aut, cre],
Juri Reimand [aut],
Jonathan Barenboim [aut]

Maintainer Helen Zhu <helen.zhu@oicr.on.ca>

Repository CRAN

Date/Publication 2019-05-04 10:10:02 UTC

R topics documented:

ActivePathways 2
brownsMethod 4
columnSignificance 5
GMT 6
hypergeometric 7

makeBackground 8
merge_p_values 8
orderedHypergeometric 9
prepareCytoscape 10

Index **11**

`ActivePathways` *ActivePathways*

Description

ActivePathways

Usage

```
ActivePathways(scores, gmt, background = NULL, geneset.filter = c(5,
  1000), cutoff = 0.1, significant = 0.05, merge.method = c("Brown",
  "Fisher", "logitp", "meanp", "sump", "sumz", "sumlog"),
  correction.method = c("holm", "fdr", "hochberg", "hommel",
  "bonferroni", "BH", "BY", "none"), return.all = FALSE,
  cytoscape.file.dir = NULL, reanalyze = FALSE)
```

Arguments

`scores`	A numerical matrix of p-values where each row is a gene and each column is a test. Rownames should be the genes and colnames the names of the tests. All values must be 0<=p<=1 with missing values removed or converted to 1
`gmt`	A GMT object to be used for enrichment analysis. If a filename, a GMT object will be read from the file
`background`	A character vector of gene names to be used as a statistical background. By default, the background is all genes that appear in `gmt`
`geneset.filter`	A numeric vector of length two giving the lower and upper limits for the size of the annotated geneset to pathways in gmt. Pathways with a geneset shorter than `geneset.filter[1]` or longer than `geneset.filter[2]` will be removed. Set either value to NA to to not enforce a minimum or maximum value, or set `geneset.filter` to `NULL` to skip filtering
`cutoff`	A maximum p-value for a gene to be used for enrichment analysis. Any genes with `adjusted.p.val > significant` will be discarded before testing
`significant`	A number in [0,1] denoting the maximum p-value for a pathway to be considered significantly enriched.
`merge.method`	Method to merge p-values. See section Merging p Values
`correction.method`	Method to correct p-values. See `p.adjust` for details
`return.all`	Whether to return results for all terms or only significant terms

`cytoscape.file.dir`
: the directory to which the output files should be written, if unspecified, files will be written to a directory called "ActivePathways.cytoscape.files". If directory does not exist, it will be automatically created. If NULL, no output files will be written.

`reanalyze`
: a boolean indicating whether the dataset will be reanalyzed with parameter or data changes. If TRUE, ActivePathways will write subdirectories in the format of Version1A, Version1B, etc to indicate sequential analyses. If FALSE, ActivePathways will write output files to the indicated directory. `reanalyze` will only be active if `cytoscape.file.dir` is not NULL.

Value

A data.table of terms containing the following columns:

term.id The id of the term

term.name The full name of the term

adjusted.p.val The associated p-value, adjusted for multiple testing

term.size The number of genes annotated to the term

overlap A character vector of the genes that overlap between the term and the query

evidence Columns of `scores` that contributed individually to enrichment of the pathway. Each column is evaluated separately for enrichments and added to the evidence field if the pathway is found.

If `return.all == FALSE` then only terms with `adjusted.p.val <= significant` will be returned, otherwise all terms will be returned.

Merging p Values

In order to obtain a single score for each gene, the p-values in `scores` are merged row-wise. There are multiple methods available that can be used to obtain this merged score. The main methods are:

Fisher or sumlog Fisher's method assumes p-values are uniformly distributed and performs a chi-squared test on the statistic sum(-2 log(p)). This method is most appropriate when the columns in `scores` are independent.

Brown Brown's method extends Fisher's method by accounting for the covariance in the columns of `scores`. It is more appropriate when the tests of significance used to create the columns in `scores` are not necessarily independent.

Other methods are also available. See `metap-package` for more details

Cytoscape

ActivePathways will write four files that can be used to build a network using Cytoscape and the EnrichmentMap and enhancedGraphics apps. The four files written are:

pathways.txt A list of significant terms and the associated p-value. Only terms with `adjusted.p.val <= significant` are written to this file

subgroups.txt A matrix indicating whether the significant pathways are found to be significant when considering only one column from `scores`. A 1 indicates that that term is significant using only that column to test for enrichment analysis

pathways.gmt A Shortened version of the supplied gmt file, containing only the terms in pathways.txt

legend.pdf A legend with colours matching contributions from columns in `scores`

How to use: Create an enrichment map in Cytoscape with the file of terms (pathways.txt) and the shortened gmt file (pathways.gmt). Upload (File > import > table > file) the subgroups file (subgroups.txt) as a table. Under the 'style' panel, set image/Chart1 to use the column 'instruct' and the passthrough mapping type. Use (legend.pdf) as a reference in final figure.

Examples

```
dat <- as.matrix(read.table(system.file('extdata',
                                        'Adenocarcinoma_scores_subset.tsv',
                                         package='ActivePathways'),
                 header=TRUE,
                 row.names='Gene'))
dat[is.na(dat)] <- 1
ActivePathways(dat,
            system.file('extdata', 'hsapiens_REAC_subset.gmt', package='ActivePathways'),
               return.all=TRUE,
               cytoscape.file.dir=NULL)
```

`brownsMethod` *Merge p-values using Brown's method*

Description

Merge p-values using Brown's method

Usage

```
brownsMethod(p.values, data.matrix = NULL, cov.matrix = NULL)
```

Arguments

`p.values`	A vector of m p-values
`data.matrix`	An m x n matrix representing m tests and n samples
`cov.matrix`	A pre-calculated covariance matrix of data.matrix. More efficient when making many calls with the same data.matrix. Only one of data.matrix and cov.matrix must be given. If both are supplied, data.matrix is ignored

Value

a p-value

`columnSignificance` *Determine which pathways are found to be significant using each column individually*

Description

Determine which pathways are found to be significant using each column individually

Usage

```
columnSignificance(scores, gmt, background, cutoff, significant,
  correction.method, pvals)
```

Arguments

`scores`	A numerical matrix of p-values where each row is a gene and each column is a test. Rownames should be the genes and colnames the names of the tests. All values must be 0<=p<=1 with missing values removed or converted to 1
`gmt`	A GMT object to be used for enrichment analysis. If a filename, a GMT object will be read from the file
`background`	A character vector of gene names to be used as a statistical background. By default, the background is all genes that appear in `gmt`
`cutoff`	A maximum p-value for a gene to be used for enrichment analysis. Any genes with `adjusted.p.val > significant` will be discarded before testing
`significant`	A number in [0,1] denoting the maximum p-value for a pathway to be considered significantly enriched.
`correction.method`	Method to correct p-values. See `p.adjust` for details
`pvals`	p-value for the pathways calculated by ActivePathways

Value

a data.table with columns 'term.id' and a column for each column in `scores`, indicating whether each pathway was found to be significant(TRUE) or not(FALSE) when considering only that column

GMT *Read and Write GMT files*

Description

Functions to read and Write Gene Matrix Transposed (GMT) files and to test if an object inherits from GMT

Usage

```
read.GMT(filename)

write.GMT(gmt, filename, path = ".")

## S3 method for class 'GMT'
x[i]

## S3 method for class 'GMT'
x[[i]]

## S3 method for class 'GMT'
x$i

is.GMT(x)
```

Arguments

filename	name of the GMT file
gmt	a GMT object
path	location of the GMT file
x	object to test
i	index of GMT object

Format

A GMT object is a named list of terms, where each term is a list with the items:

id The term id

name The full name of the term

genes A character vector of genes annotated to this term

Details

A GMT file describes gene sets, such as pathways. GMT files are tab delimeted and each row contains a term id, a term name, and all genes annotated to the term.

Value

read.GMT returns a GMT object.
write.GMT returns NULL.
is.GMT returns TRUE if x is a GMT object, else FALSE

Examples

```
gmt <- read.GMT(system.file('extdata', 'hsapiens_REAC_subset.gmt', package='ActivePathways'))
is.GMT(gmt)
gmt[1:10]
gmt[[1]]
gmt[1]$id
gmt[1]$genes
gmt[1]$name
gmt$`REAC:3108214`

  write.GMT(gmt, 'filename.gmt', path = tempdir())
```

---

hypergeometric *Hypergemoetric Test*

Description

Perform a Hypergeometric test, aka Fisher's exact test, on a 2x2 contingency table with a 'greater' alternative hypothesis. That is, find the probability that counts[1, 1] or more genes would be found in annotations, assuming the null hypothesis.

Usage

```
hypergeometric(counts)
```

Arguments

counts	a 2x2 numerical matrix representing a contingency table

Value

a p-value

makeBackground *Make a background list of genes*

Description

Returns a character vector of all genes in a GMT object

Usage

```
makeBackground(gmt)
```

Arguments

gmt	a GMT object

Value

a character vector containing all genes in GMT

Examples

```
gmt <- read.GMT(system.file('extdata', 'hsapiens_REAC_subset.gmt', package='ActivePathways'))
makeBackground(gmt)
```

merge_p_values *Merge a list or matrix of p-values*

Description

Merge a list or matrix of p-values

Usage

```
merge_p_values(scores, method = c("Fisher", "Brown", "logitp", "meanp",
  "sump", "sumz", "sumlog"))
```

Arguments

scores	Either a list of p-values or a matrix where each column is a test
method	Method to merge p-values. See 'methods' section below.

Value

If scores is a vector or list, returns a number. If scores is a matrix, returns a named list of p-values merged by row

Methods

There are multiple methods available that can be used to merge a list of p-values. The main methods encouraged by this package are:

Fisher or sumlog Fisher's method assumes p-values are uniformly distributed and performs a chi-squared test on the statistic sum(-2 log(p)). This method is most appropriate when the columns in `scores` are independent.

Brown Brown's method extends Fisher's method by accounting for the covariance in the columns of `scores`. It is more appropriate when the tests of significance used to create the columns in `scores` are not necessarily independent. Note that the "Brown" method cannot be used with a single list of p-values. However, in this case Brown's method is identical to Fisher's method.

Other methods are also available. See `metap-package` for more details

Examples

```
merge_p_values(c(0.05, 0.09, 0.01))
merge_p_values(list(a=0.01, b=1, c=0.0015, d=0.025), method='meanp')
merge_p_values(matrix(data=c(0.03, 0.061, 0.48, 0.052), nrow=2), method='Brown')
```

`orderedHypergeometric` *Ordered Hypergeometric Test*

Description

Perform a Hypergeometric, aka Fisher's Exact test, on a list of genes ordered by significance against a list of annotation genes

Usage

```
orderedHypergeometric(genelist, background, annotations)
```

Arguments

`genelist`	character vector of gene names. List of differentially expressed genes being tested for enrichment
`background`	character vector of gene names. List of all genes being used as a statistical background
`annotations`	character vector of gene names. List of genes annotated to the term being tested.

Details

The hypergeometric test is run with increasingly large numbers of genes starting from the top, and the lowest p-value is returned

Value

a list with the items:

p.val The lowest obtained p-value

ind The index of `genelist` such that `genelist[1:ind]` gives the lowest p-value

`prepareCytoscape` *Prepare files for building an Enrichment Map in Cytoscape*

Description

This function writes four files that are used to build an network using Cytoscape and the EnrichmentMap app. The four files written are:

pathways.txt A list of significant terms and the associated p-value. Only terms with `adjusted.p.val <= significant` are written to this file

subgroups.txt A matrix indicating whether the significant pathways are found to be significant when considering only one column from `scores`. A 1 indicates that that term is significant using only that column to test for enrichment analysis

pathways.gmt A Shortened version of the supplied gmt file, containing only the terms in pathways.txt

legend.pdf A legend with colours matching contributions from columns in `scores`

Usage

```
prepareCytoscape(terms, gmt, file_dir, col.significance)
```

Arguments

`terms`	a data.table with columns 'term.id', 'term.name', 'adjusted.p.val'
`gmt`	an abridged gmt object containing only the pathways that were found to be significant
`file_dir`	user defined directory to write output files
`col.significance`	a data.table with a column 'term.id' and a column for each test indicating whether a pathway is signficiant (TRUE) or not (FALSE) when considering only that column. If contribution==TRUE, use col.significance=NULL and this will be skipped

Value

None

Index

[.GMT *(GMT)*, 6
[[.GMT *(GMT)*, 6
$.GMT *(GMT)*, 6

ActivePathways, 2

brownsMethod, 4

columnSignificance, 5

GMT, 6, *8*
gmt *(GMT)*, 6

hypergeometric, 7

is.GMT *(GMT)*, 6

makeBackground, 8
merge_p_values, 8

orderedHypergeometric, 9

p.adjust, *2*, *5*
prepareCytoscape, 10

read.GMT *(GMT)*, 6

write.GMT *(GMT)*, 6